RAINFALL PREDICTION USING LSTM DEEP LEARNING MODEL

OR 610 PROJECT

BY TUAN NGUYEN

MAY 2$^{\text{ND}}$, 2023

CONTENT

1. INTRODUCTION

RAINFALL

Rainfall (or Precipitation)

- Condensed atmospheric water vapor pulled by gravity

- Affects numerous human activities

- Can cause hazardous weather conditions

→ Knowing **when** it will rain in advance can lead better mitigation plan/take advantage of these events

CURRENT METHODS

Physical Based – Simulation Models

- NOAA's Global Forecast System (GFS)
- ECMWF's Integrated Forecasting System (IFS)

Problem: Costly To Operate

Statistical Based – Statistic/ML/DL Models

- ARIMA, SVM, MLP, ANFIS, RNN/LSTM
  (Historical Records)
- CNN, ConvLSTM
  (Radar Echo, Satellite Images)

Problem: Slower To Adapt To New Data

CHALLENGES – LONG TERM PREDICTION:

- Many influencing factors
- "Butterfly Effect"
- Non-deterministic
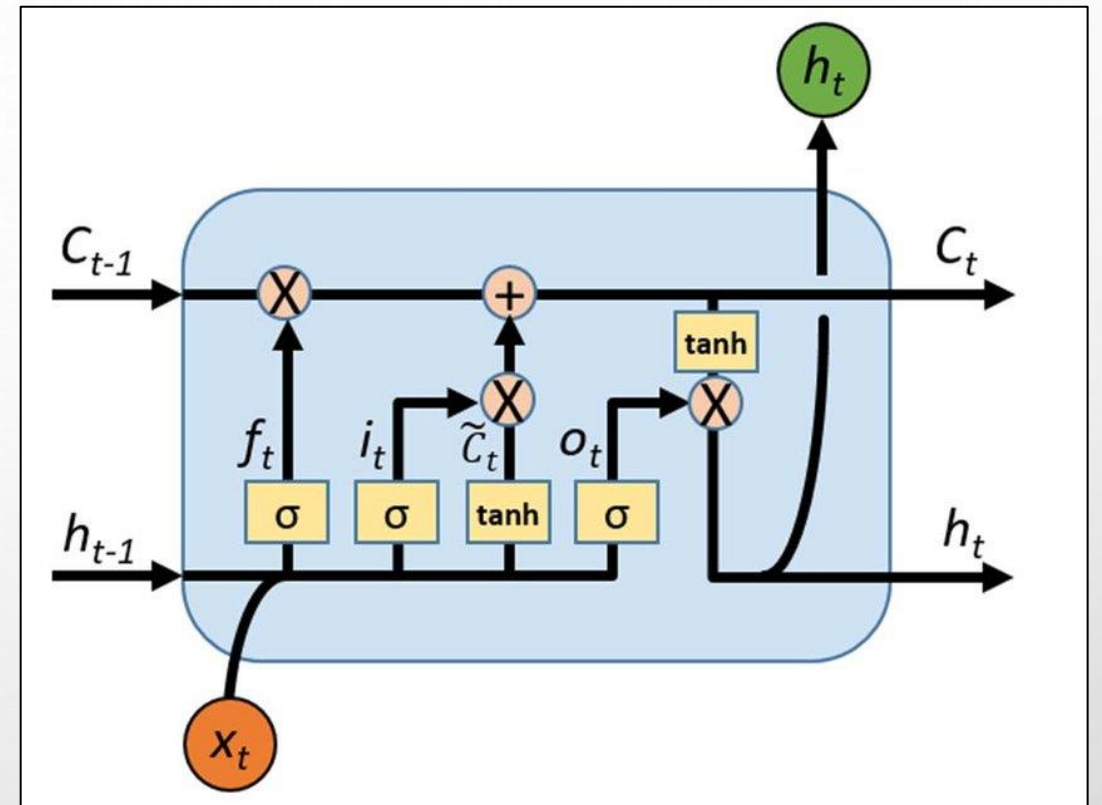- Amplified error per time step

POSSIBLE SOLUTIONS:

- Predict only Monthly/Weekly average

→ Lose temporal information

SOLUTION PROPOSAL

- LSTM Model Architecture

- Predict Daily Rainfall In Millimeters (Mm)

- Input: 30 – 90 Days Prior Weather Data

- Output:
  - Daily Iterative Model: 1-day Prediction
  - Single Prediction Model: 30-days Prediction

2. DATASET

ORIGINAL DATASET

- "Rain in Australia" dataset

- Kaggle dataset, 145,460 observations of 23 columns

- Daily measurements from 49 Stations:

  - Temperature (Max, Min)
  - Temperature (9am, 3pm)
  - Rainfall
  - Evaporation
  - Sunshine hours
  - Wind gust direction (9am, 3pm)

  - Wind gust speed (9am, 3pm)
  - Humidity (9am, 3pm)
  - Pressure (9am, 3pm)
  - Cloud (9am, 3pm)

Date	Day	Temps Min °C	Temps Max °C	Rain mm	Evap mm	Sun hours	Max wind gust Dir	Spd km/h	Time local	9 am Temp °C	RH %	Cld 8th	Dir	Spd km/h	MSLP hPa	3 pm Temp °C	RH %	Cld 8th	Dir	Spd km/h	MSLP hPa
1	Tu	18.2	24.0	0.2	4.6	9.5	WNW	69	06:50	19.2	45	2	WNW	35	992.9	23.1	30	2	WNW	31	992.0
2	We	11.1	20.5	0.6	13.0	12.8	W	67	12:54	14.0	44	1	W	31	1003.2	19.7	29	1	SW	22	1005.9
3	Th	11.1	22.0	0	7.8	8.9	W	56	07:41	15.9	50	1	WSW	31	1014.7	20.0	42	7	SE	20	1016.6
4	Fr	13.4	23.1	1.0	6.0	5.7	SSE	26	23:24	15.8	82	6	E	6	1026.9	22.8	55	5	ENE	17	1026.8
5	Sa	13.4	23.6	0.2	4.4	11.8	ENE	37	15:20	17.7	76	3	WSW	4	1029.7	22.2	60	2	ENE	24	1026.9
6	Su	13.3	24.0	0	4.0	12.1	ENE	39	16:12	19.0	75	1	ESE	6	1026.6	23.1	59	1	E	26	1023.3
7	Mo	15.4	24.2	0	9.8	12.3	NE	41	11:44	20.4	80	5	E	15	1023.4	23.4	58	2	ENE	24	1020.6
8	Tu	16.0	24.2	1.2	8.0	11.0	ENE	35	15:55	21.1	67	4	ENE	15	1021.5	24.0	59	2	ENE	24	1020.2
9	We	14.9	24.2	0.2	8.0	10.3	E	33	12:37	19.2	77	7	WNW	6	1023.0	23.6	46	1	ENE	26	1021.9
10	Th	14.9	24.4	0	7.8	9.3	ENE	43	17:10	17.7	76	7	WNW	7	1023.1	24.1	55	1	ENE	24	1019.1
11	Fr	14.9	24.9	0	7.8	9.1	ENE	39	12:12	20.1	60	4	ENE	7	1015.2	24.4	56	7	NE	17	1011.4
12	Sa	16.0	27.9	0	7.8	9.5	ESE	26	13:53	20.7	77	1	N	4	1012.1	27.7	51	3	ESE	19	1011.1
13	Su	18.9	25.6	0	6.4	1.7	NNW	46	16:21	21.2	85	8	SE	2	1010.2	25.2	70	7	NNE	24	1005.0
14	Mo	18.1	27.9	37.6	4.2	10.8	W	67	13:20	20.6	76	1	WNW	11	1001.3	27.6	34	2	W	33	1000.5
15	Tu	16.2	24.9	0.2	9.6	9.2	SW	33	07:39	20.4	54	2	WSW	19	1008.4	22.6	62	7	E	20	1005.6
16	We	13.0	21.6	0.8	8.0	9.1	S	54	14:56	15.9	42	2	WSW	22	1010.1	15.3	56	7	WSW	24	1009.5
17	Th	10.0	20.3	3.4	5.2	12.6	SSE	46	14:00	15.4	62	1	W	17	1015.1	19.1	45	1	SSE	24	1013.7
18	Fr	12.4	22.2	0	8.0	12.9	SSE	37	10:05	17.1	52	1	SW	17	1016.4	21.2	47	1	SSE	24	1014.0
19	Sa	12.9	23.9	0	7.6	12.3	ENE	50	15:05	19.0	69	3	NNW	13	1012.3	22.4	64	3	NE	30	1005.9
20	Su	18.0	27.4	0.8	7.8	13.3	W	70	10:36	23.9	56	1	WNW	19	995.9	25.3	30	3	WNW	37	998.9
21	Mo	17.1	24.5	0	12.2	12.7	W	83	13:27	21.8	35	1	WNW	24	1002.3	22.3	31	2	WNW	43	1002.7
22	Tu	11.7	24.0	0	12.0	13.0	W	63	06:50	15.3	42	1	WSW	37	1012.8	23.7	28	1	W	30	1011.1
23	We	13.7	28.3	0	8.0	12.4	W	54	08:27	18.8	48	1	W	28	1014.2	27.2	31	2	WNW	19	1012.5
24	Th	15.8	25.4	0	11.2	12.0	E	33	15:17	20.7	49	1	W	13	1018.4	25.1	48	1	E	19	1016.7
25	Fr	16.6	24.1	0	7.8	12.7	E	31	11:28	20.0	77	3	NNE	7	1018.4	23.5	67	2	E	22	1015.3
26	Sa	17.0	23.3	0.6	6.2	11.1	SSE	41	07:03	19.1	70	5	SSW	15	1015.9	23.0	48	2	E	19	1012.5
27	Su	17.0	26.8	0	8.0	3.7	WSW	54	18:51	19.7	74	7	NNE	6	1004.9	24.8	68	4	E	11	1001.1
28	Mo	17.1	23.8	8.8	3.8	8.9	SSW	31	03:28	20.1	82	6	SSW	19	1006.8	22.7	63	6	ESE	22	1005.5
29	Tu	15.9	25.4	0	4.0	12.5	SSE	46	11:55	20.4	46	1	SW	20	1011.6	23.6	52	1	SE	30	1013.1
30	We	16.0	22.9	0	11.4	1.3	E	31	14:16	18.8	59	7	E	13	1019.4	20.7	56	7	E	24	1016.9

Reference: http://www.bom.gov.au/climate/dwo/IDCJDW2124.latest.shtml

MISSING VALUES

Overall Missing Value in each column:

- Date: 0.00 %
- Location: 0.00 %
- MinTemp: 1.02 %
- MaxTemp: 0.86 %
- Rainfall: 2.24 %
- Evaporation: 43.16%
- Sunshine: 48.01%
- WindGustDir: 7.09 %

- WindGustSpeed: 7.05 %
- WindDir9am: 7.26 %
- WindDir3pm: 2.90 %
- WindSpeed9am: 1.21 %
- WindSpeed3pm: 2.105%
- Humidity9am: 1.82 %
- Humidity3pm: 3.098%
- Pressure9am: 10.35%

- Pressure3pm: 10.33%
- Cloud9am: 38.42%
- Cloud3pm: 40.81%
- Temp9am: 1.21 %
- Temp3pm: 2.48 %
- RainToday: 2.24 %
- RainTomorrow: 2.24 %

|	Date	Location	MinTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustDir	WindGustSpeed	WindDir9am	WindDir3pm	WindSpeed9am	WindSpeed3pm	Humidity9am	Humidity3pm	Pressure9am	Pressure3pm	Cloud9am	Cloud3pm	Temp9am	Temp3pm	RainToday	RainTomorrow
Albury	0.0	0.0	0.4	0.4	1.0	100.0	100.0	1.1	1.1	17.3	1.5	0.3	0.4	0.3	0.3	0.2	0.3	57.6	53.1	0.3	0.3	1.0	1.0
BadgerysCreek	0.0	0.0	1.2	1.0	2.7	100.0	100.0	2.5	2.5	14.5	2.8	1.6	1.6	1.8	1.8	5.6	5.7	100.0	100.0	1.4	1.5	2.7	2.7
Cobar	0.0	0.0	0.2	0.1	0.7	11.4	81.7	1.4	1.4	0.9	0.5	0.6	0.4	4.2	4.2	1.0	0.9	12.6	10.2	0.3	0.2	0.7	0.7
CoffsHarbour	0.0	0.0	0.5	0.6	1.9	40.7	50.3	11.9	11.9	11.5	10.1	9.8	9.8	0.6	0.7	10.3	10.4	27.0	25.9	0.5	0.5	1.9	1.9
Moree	0.0	0.0	0.1	0.0	5.2	23.8	31.7	2.6	2.6	3.6	1.8	1.4	1.4	0.1	0.2	0.0	0.0	17.9	13.4	0.0	0.0	5.2	5.2
Newcastle	0.0	0.0	11.4	7.7	2.8	100.0	100.0	100.0	100.0	46.8	43.4	8.4	31.2	7.9	30.9	100.0	100.0	7.9	30.7	7.7	30.7	2.8	2.8
NorahHead	0.0	0.0	1.0	1.0	2.5	100.0	100.0	1.8	1.8	3.6	1.6	1.3	1.1	1.3	1.1	1.1	0.9	100.0	100.0	1.3	1.0	2.5	2.5
NorfolkIsland	0.0	0.0	0.0	0.0	1.5	7.8	14.6	1.5	1.5	1.0	0.2	0.3	0.1	0.2	0.1	0.3	0.1	2.2	2.0	0.2	0.1	1.5	1.5
Penrith	0.0	0.0	0.9	0.8	2.5	100.0	100.0	2.1	2.1	18.4	2.4	1.4	0.7	3.2	2.3	100.0	100.0	100.0	100.0	1.3	0.7	2.5	2.5
Richmond	0.0	0.0	0.7	0.5	1.9	46.0	100.0	1.9	1.8	28.3	3.9	1.4	1.2	2.0	2.0	1.0	1.2	79.9	80.7	1.3	1.4	1.9	1.9
Sydney	0.0	0.0	0.1	0.1	0.2	1.5	0.5	31.0	31.0	1.7	1.0	0.8	0.7	0.4	0.4	0.6	0.6	17.0	16.8	0.1	0.1	0.2	0.2
SydneyAirport	0.0	0.0	0.0	0.0	0.1	1.5	0.5	1.7	1.7	0.3	0.1	0.1	0.0	0.1	0.1	0.3	0.1	0.1	0.0	0.0	0.1	0.1
WaggaWagga	0.0	0.0	0.0	0.0	1.1	4.6	14.4	0.9	0.9	4.0	0.5	0.0	0.0	0.1	0.1	0.0	0.0	7.4	5.7	0.0	0.0	1.1	1.1
Williamtown	0.0	0.0	0.1	0.1	15.2	35.3	55.0	4.5	4.5	2.9	0.6	0.6	0.3	0.3	0.2	0.4	0.4	20.0	21.6	0.2	0.1	15.2	15.2
Wollongong	0.0	0.0	0.5	0.4	1.9	100.0	100.0	1.6	1.6	4.1	1.3	1.1	1.0	1.0	1.0	1.6	0.6	55.8	53.4	1.0	1.0	1.9	1.9
Canberra	0.0	0.0	0.2	0.1	0.5	46.7	55.7	9.9	9.8	15.7	6.6	6.7	6.5	1.8	0.3	6.6	6.4	31.2	36.8	0.5	0.2	0.5	0.5
Tuggeranong	0.0	0.0	0.0	0.1	1.3	100.0	100.0	1.4	1.4	21.2	1.5	0.8	0.7	0.7	0.4	0.5	0.4	100.0	100.0	0.7	0.4	1.3	1.3
MountGinini	0.0	0.0	3.0	1.7	4.4	100.0	100.0	10.5	10.5	8.2	7.6	7.2	7.1	14.3	10.2	100.0	100.0	100.0	100.0	13.8	9.7	4.4	4.4
Ballarat	0.0	0.0	0.0	0.0	0.4	100.0	100.0	1.0	1.0	3.0	0.6	0.3	0.1	0.7	0.2	0.8	0.9	18.0	34.5	0.4	0.2	0.4	0.4
Bendigo	0.0	0.0	0.1	0.2	0.2	61.0	100.0	1.3	1.3	8.0	1.4	0.1	0.1	0.2	0.1	0.5	0.3	31.8	25.4	0.1	0.1	0.2	0.2
Sale	0.0	0.0	0.0	0.0	0.3	39.0	39.6	4.7	4.7	2.7	0.5	0.1	0.2	0.6	0.5	0.5	0.5	11.9	13.0	0.1	0.0	0.3	0.3
MelbourneAirport	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.5	0.5	1.3	0.4	0.0	0.0	0.3	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Melbourne	0.0	0.0	15.0	15.1	23.7	0.1	0.0	0.4	0.4	1.6	0.4	0.1	0.0	15.1	15.3	15.0	15.1	32.4	34.6	15.1	15.2	23.7	23.7
Mildura	0.0	0.0	0.0	0.0	0.1	3.9	4.4	0.2	0.2	3.1	0.3	0.1	0.0	0.1	0.1	0.1	0.1	2.8	4.9	0.0	0.0	0.1	0.1
Nhil	0.0	0.0	0.3	0.4	0.6	100.0	100.0	0.7	0.7	3.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	100.0	100.0	0.3	0.3	0.6	0.6
Portland	0.0	0.0	0.3	0.0	0.4	13.6	14.7	1.2	1.2	2.3	0.2	0.1	0.1	1.6	0.4	0.4	0.3	13.7	17.2	1.6	0.4	0.4	0.4
Watsonia	0.0	0.0	0.2	0.0	0.3	0.1	0.0	1.2	1.2	6.3	1.2	0.0	0.0	0.5	0.0	0.1	0.0	0.0	0.0	0.5	0.0	0.3	0.3
Dartmoor	0.0	0.0	2.3	2.1	2.2	13.6	14.7	2.3	2.3	18.9	3.3	2.2	2.1	7.6	7.5	2.1	2.2	100.0	100.0	2.1	2.1	2.2	2.2
Brisbane	0.0	0.0	0.3	0.4	1.0	0.6	1.5	1.3	1.3	2.2	1.1	0.0	0.3	0.1	0.5	0.0	0.3	0.0	0.1	0.1	0.5	1.0	1.0
Cairns	0.0	0.0	0.0	0.0	1.7	12.2	15.7	0.6	0.6	1.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	6.8	7.6	0.0	0.0	1.7	1.7
GoldCoast	0.0	0.0	0.1	0.2	2.0	100.0	100.0	3.9	3.9	2.1	1.2	1.3	1.0	0.5	0.4	0.0	0.0	100.0	100.0	0.0	2.0	2.0	2.0
Townsville	0.0	0.0	0.1	0.0	0.2	3.0	13.9	0.8	0.8	4.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	5.7	9.1	0.0	0.0	0.2	0.2
Adelaide	0.0	0.0	0.1	0.1	3.2	46.6	44.6	0.8	0.8	8.6	0.5	0.2	0.2	0.2	0.2	0.3	0.2	100.0	100.0	0.2	0.1	3.2	3.2
MountGambier	0.0	0.0	0.1	0.2	0.3	13.5	14.6	1.7	1.7	2.8	0.6	0.3	0.3	0.4	0.3	0.1	0.2	3.6	3.6	0.1	0.2	0.3	0.3
Nuriootpa	0.0	0.0	0.4	0.3	0.2	4.1	5.4	1.6	1.1	4.8	1.1	0.7	0.9	0.3	0.4	0.2	0.3	4.2	25.6	0.2	0.2	0.2	0.2
Woomera	0.0	0.0	0.1	0.1	0.6	14.6	33.3	1.3	1.3	1.0	0.6	0.1	0.2	0.4	0.9	0.4	0.5	12.4	21.9	0.0	0.1	0.6	0.6
Albany	0.0	0.0	2.1	1.8	0.8	9.5	17.1	100.0	100.0	8.1	25.8	2.0	24.1	1.6	23.6	0.8	0.8	1.2	23.4	1.3	23.4	0.8	0.8
Witchcliffe	0.0	0.0	0.3	0.2	1.9	100.0	100.0	1.1	1.1	8.3	0.6	0.2	0.1	8.0	7.8	4.5	4.5	100.0	100.0	0.1	0.1	1.9	1.9
PearceRAAF	0.0	0.0	0.7	0.7	8.2	100.0	0.2	4.8	4.8	3.0	0.9	0.7	0.9	0.7	0.9	0.7	0.9	34.7	37.5	0.6	0.9	8.2	8.2
PerthAirport	0.0	0.0	0.0	0.0	0.0	0.0	0.2	1.5	1.5	1.3	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0

SELECTED STATIONS

21 STATIONS SELECTED (< 30% N/A per column)

- Sydney Airport
- Sydney
- Wagga Wagga
- Moree
- Cobar
- Melbourne Airport (major)
- Mildura
- Portland
- Watsonia
- Dartmoor
- Brisbane
- Cairns
- Townsville
- Nuriootpa
- Perth Airport
- Perth
- Alice Springs
- Darwin
- Norfolk Island
- Mount Gambier
- Hobart

64,353 Observations of **23** Columns

RAINFALL DATA
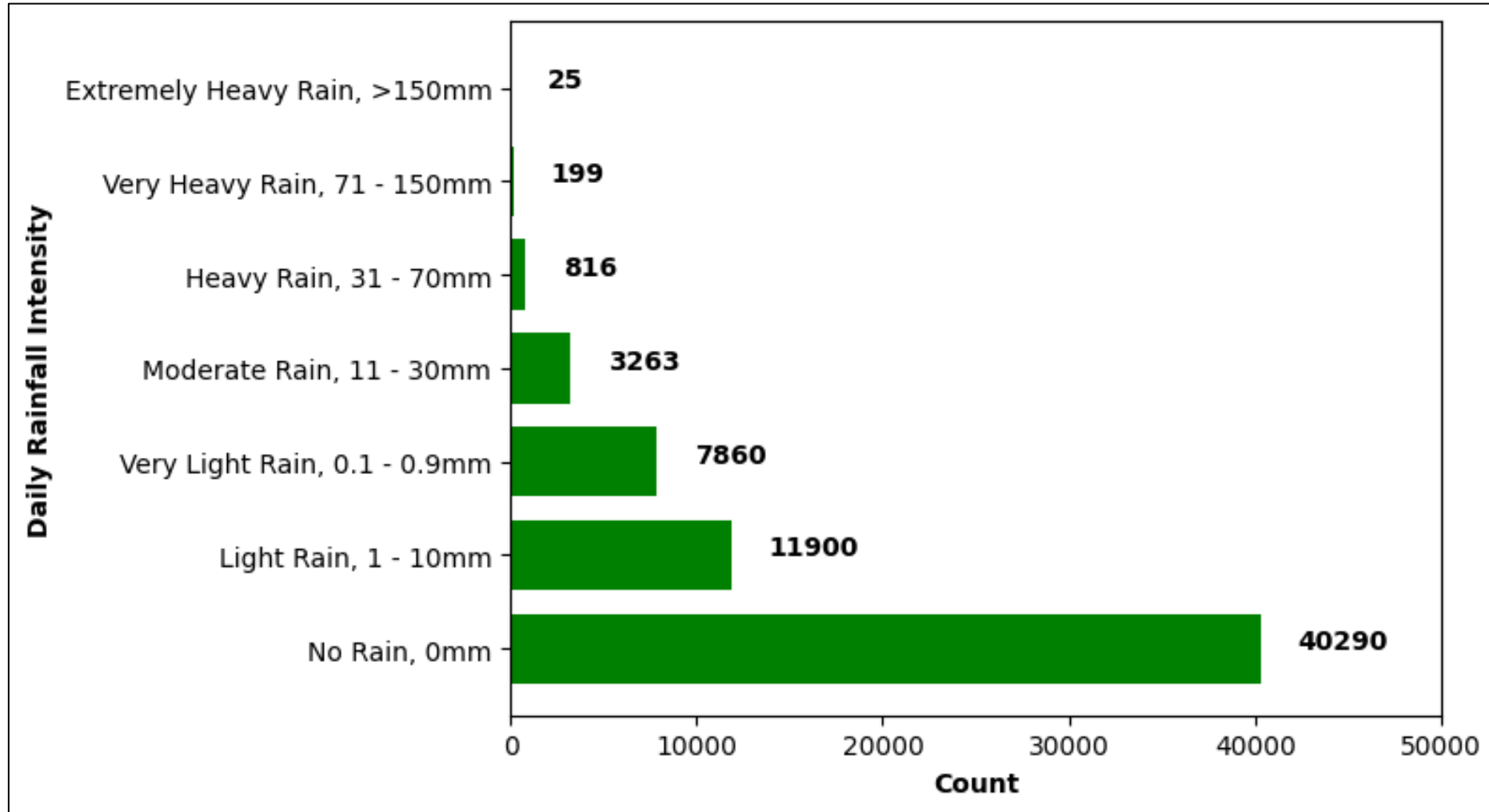
Observed Daily Rainfall Frequency by Month for all 21 Stations (2007-2017)

PREPROCESS

- Missing Values:

  - Create separate data frame for each station

  - Fill N/A using Mean & Mode at each station

  - Recombine all frames

- Rescale data (sklearn):

  - MinMaxScaler(): Rainfall

  - StandardScaler(): Temperature, Evaporation, Wind speed, Pressure, Cloud

  - Convert from % to ratio: Humidity

TRAIN / VAL / TEST SPLIT

- Create sequence with Sliding window

- Split ratio:

  - 0.8 Train

  - 0.15 Val

  - 0.05 Test

- Ingest using Torch Data Loader

- Only enabled shuffled in Train Data Loader
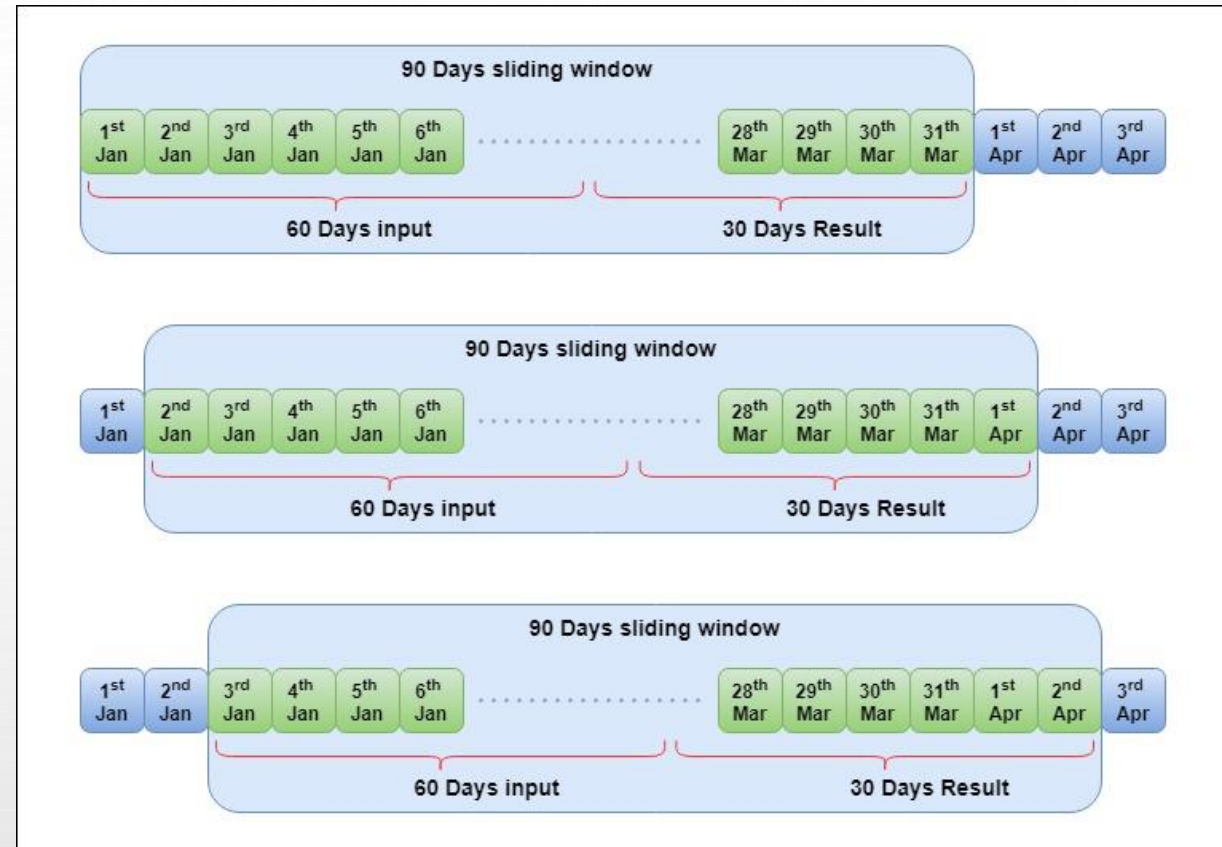
3. MODEL

MODEL ARCHITECTURE
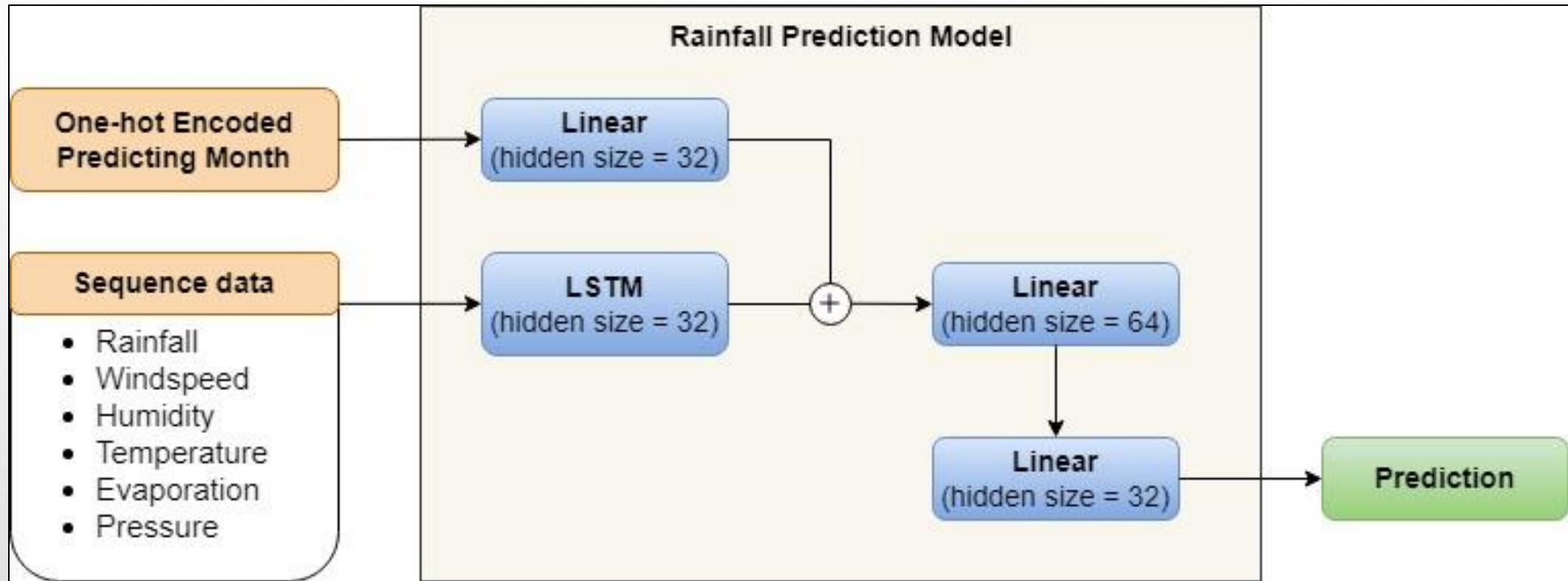
- Daily Iterative: Last linear output size = 1

- Single Prediction: Last linear output size = 30

LOSS FUNCTION

Mean Square Error (MSE):

- Sensitive to outlier

- Predicts closer to mean

Mean Absolute Error (MAE):

- Equal weights to outliers

- Predicts closer to median

Huber Loss:

- Combine advantages of MSE & MAE

- Use MSE if loss below delta

- Use MAE if loss above delta

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

$$loss = \begin{cases} \frac{1}{2} * (x - y)^2 & if\ (|x - y| \leq \delta) \\ \delta * |x - y| - \frac{1}{2} * \delta^2 & otherwise \end{cases}$$

MODEL TRAINING

- 50 epochs with Adam Optimizer (lr = 0.001)

- Different input sequence length used (30, 60, 90)

- Various batch size used (32,64,128,256)

- Best weight set (lowest Val loss) was recorded for each epoch

- Data Loader shuffle seed also affect performance

Training Loss (Daily Iterative)
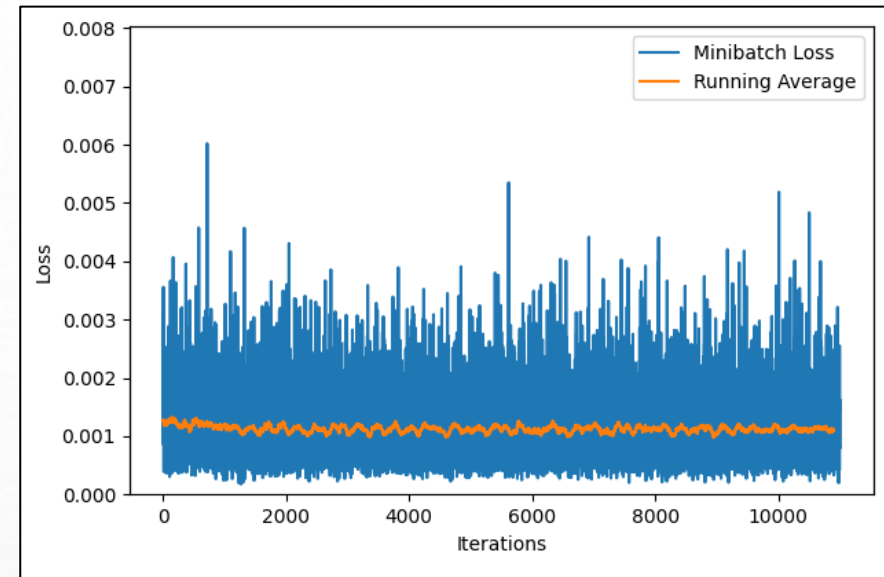
Validation Loss (Daily Iterative)
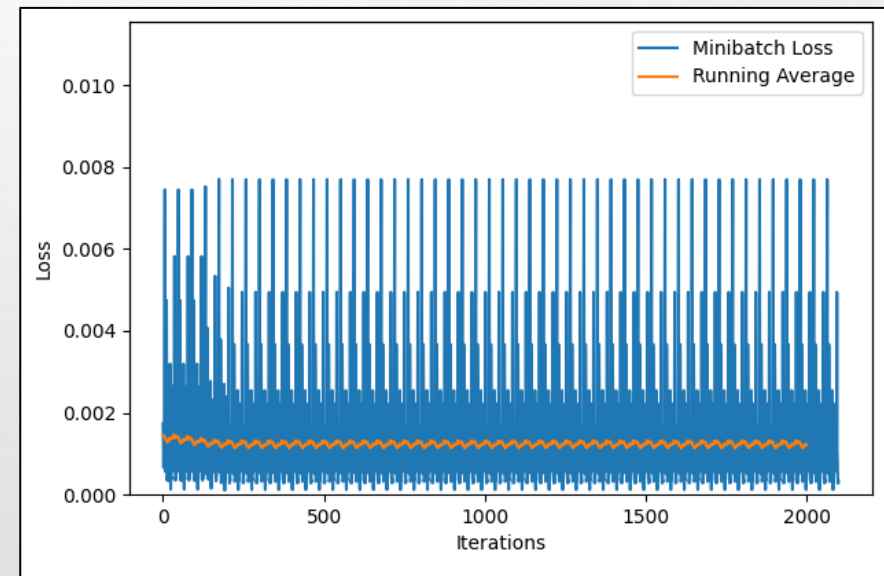
4. EXPERIMENTAL RESULTS

DAILY ITERATIVE MODEL

Pred (y-axis) versus Real (x-axis)

Batch Size	Metric	Input Sequence Length		
		30 days	60 days	90 days
32	MSE	52.81804	56.30755	48.95148
	RMSE	7.26760	7.50383	6.99653
	MAE	2.26170	2.28474	2.00663
64	MSE	61.54731	61.32821	**45.00431**
	RMSE	7.84520	7.83123	**6.70852**
	MAE	2.20892	2.25611	**2.2923**
128	MSE	51.45301	57.52034	46.68670
	RMSE	7.17307	7.58421	6.83276
	MAE	2.30552	2.64371	2.04567
256	MSE	54.02511	57.29244	**45.78643**
	RMSE	7.35017	7.56917	**6.76656**
	MAE	2.49377	2.59183	**2.20926**
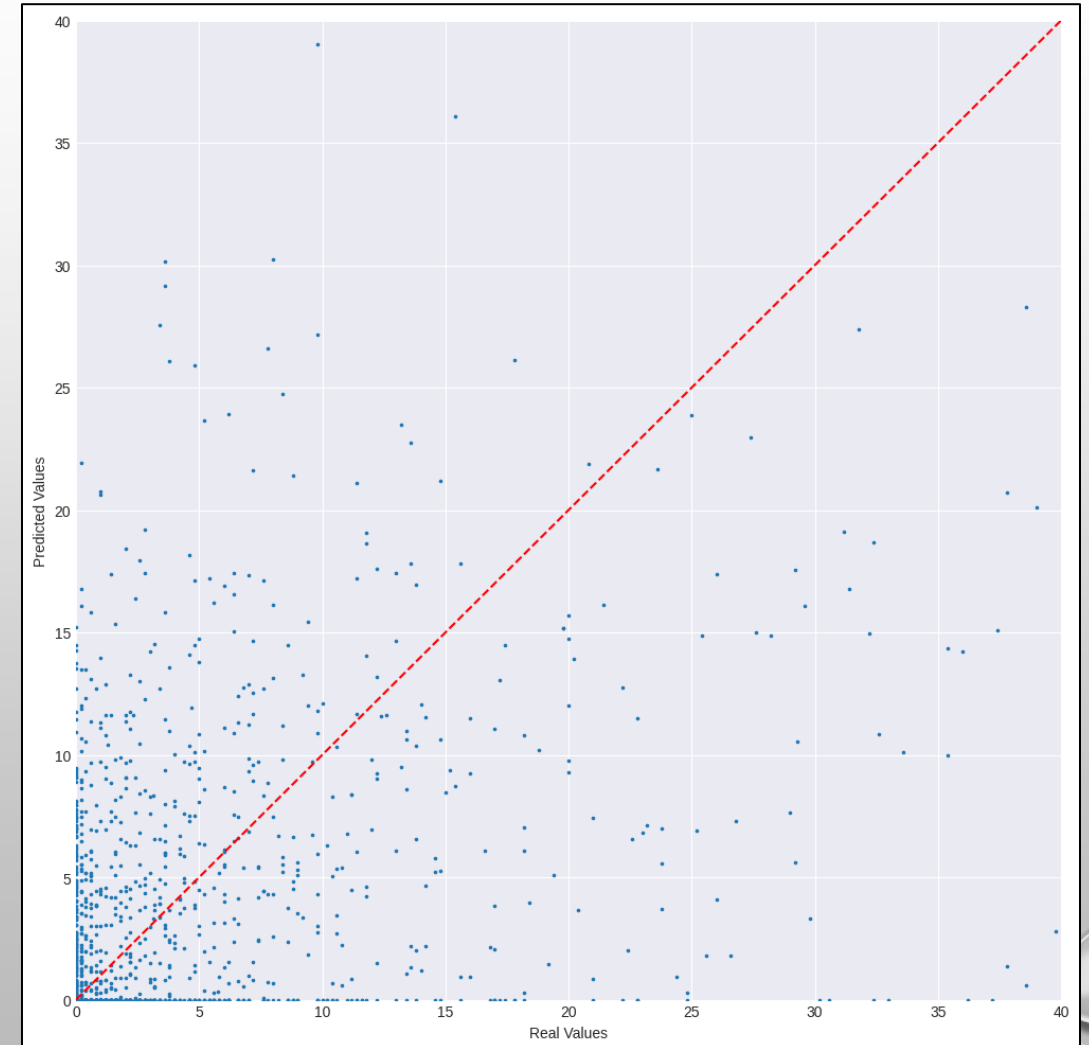
DAILY ITERATIVE MODEL

Time-Series Prediction
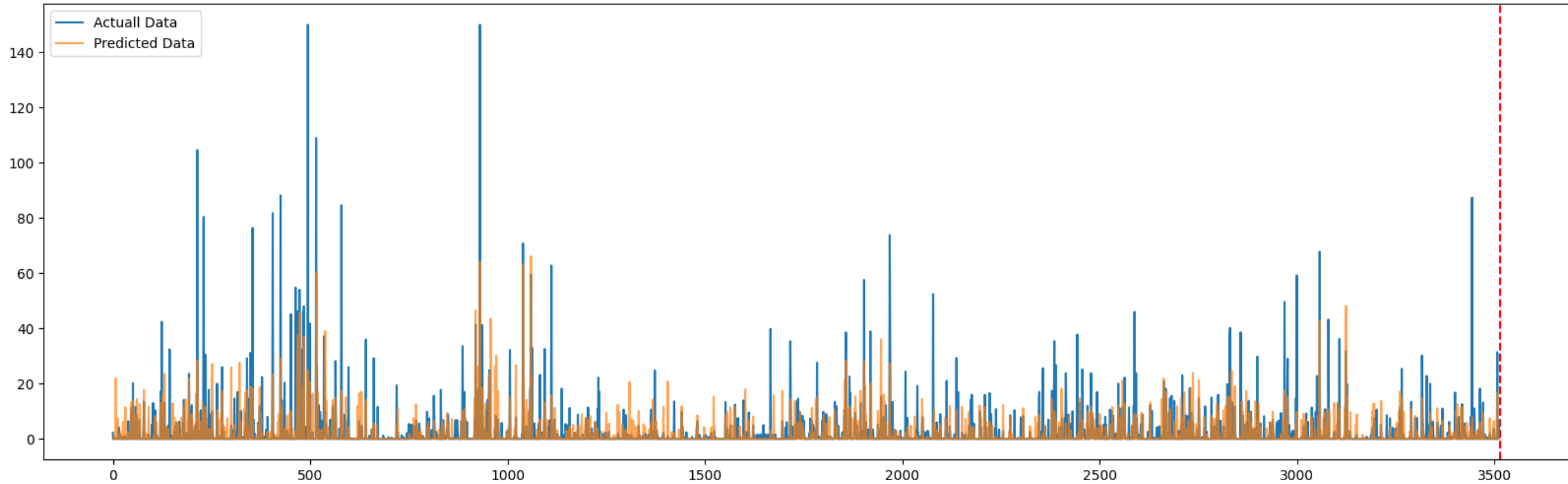
DAILY ITERATIVE MODEL

Residual Plot
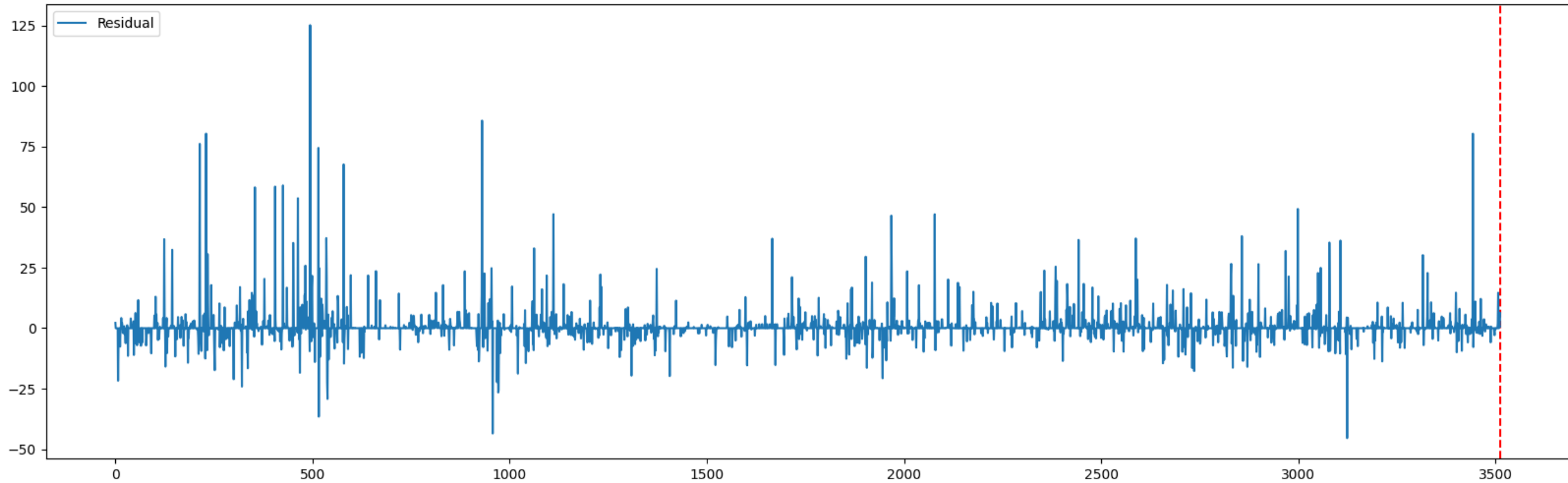
SINGLE PREDICTION MODEL

- Fixed input length of 60 days

- Output length of 30 days

- Batch size set to 128

Model performance:

- MSE = 69.4754

- RMSE = 8.33519

- MAE = 2.9877

Pred (y-axis) versus Real (x-axis)
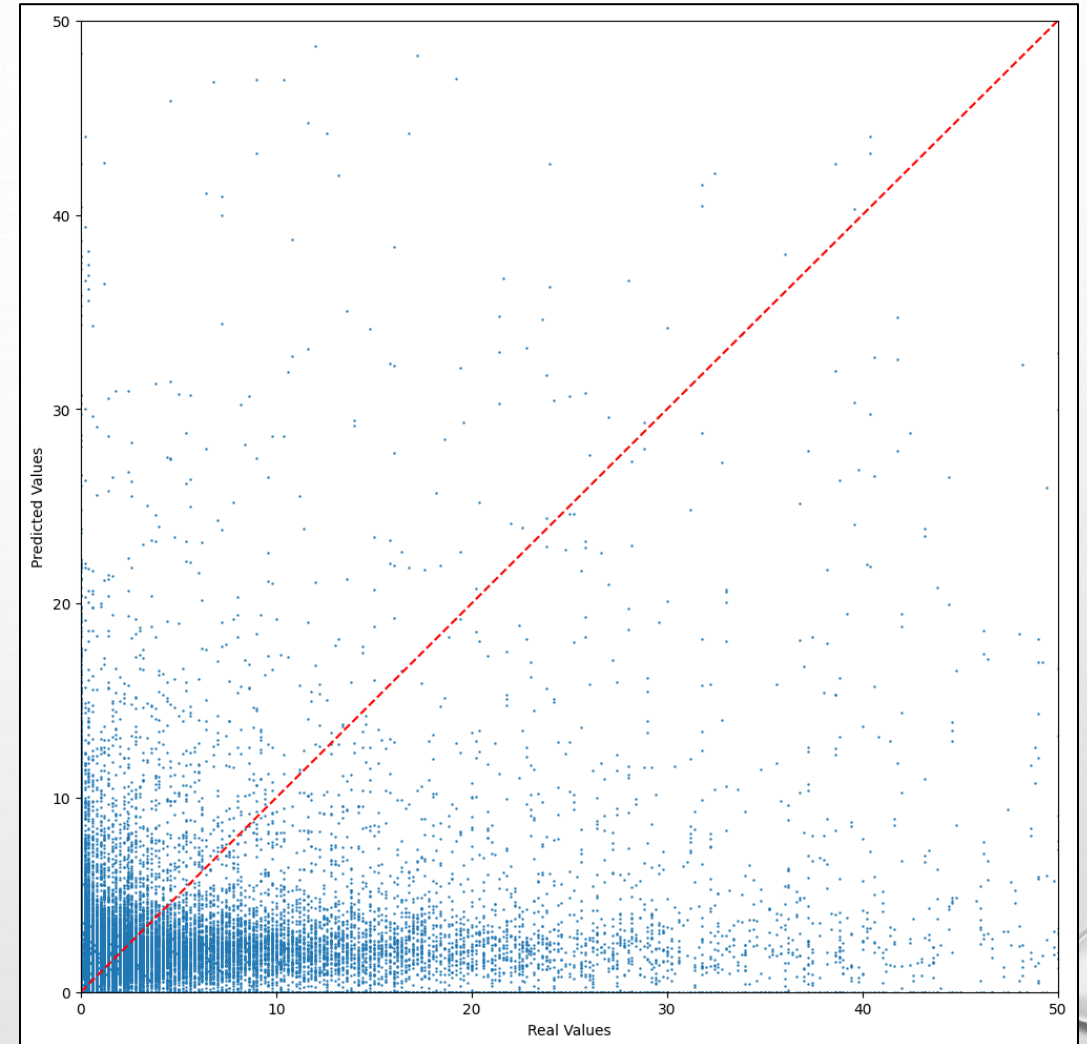
SINGLE PREDICTION MODEL

Time-Series Prediction
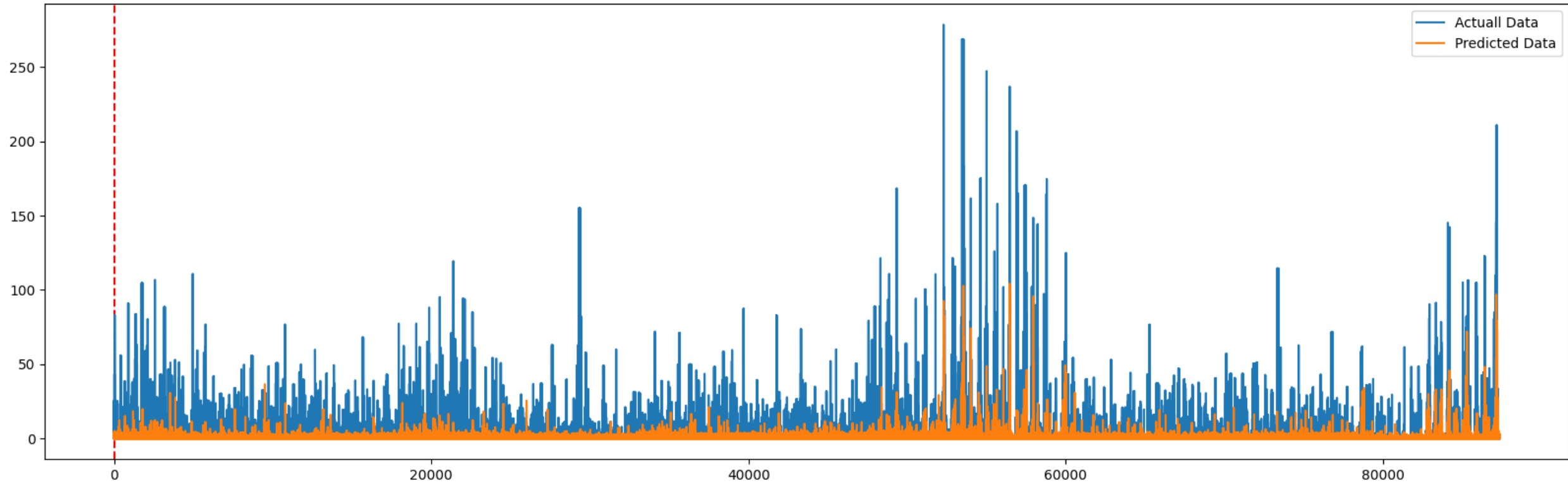
RESULT COMPARISON

Model (best)	Metric	Model Type	
		Daily Iterative	Single Prediction
Trained Model	MSE	**45.78643**	69.4754
	RMSE	**6.76656**	**8.33519**
	MAE	**2.20926**	**2.9877**
All zeroes	MSE	75.32471	78.96564
	RMSE	8.67898	**8.88626**
	MAE	2.37367	**2.30722**
Randomized	MSE	137.27216	392.0345
	RMSE	11.71632	19.79986
	MAE	9.8348	11.92328

5. DISCUSSION

EFFECT OF LOSS FN

MSE Loss (nn.MSELoss):

- Encourage model to make Non-zero predictions

MAE Loss (nn.L1Loss):

- Encourage model to make all-zero predictions

- Behavior caused by zero-inflated data

MODELS BEHAVIOR

Daily Iterative Model

- Over-predict Amount

- Under-predict Rainfall Event occurrences
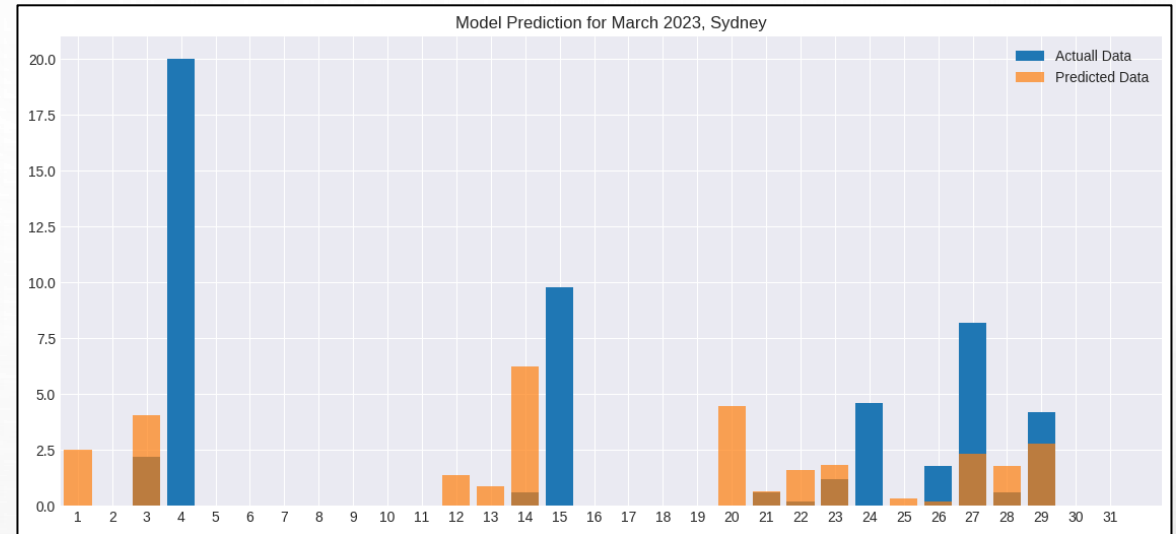
Single Prediction Model:

- Under-predict Amount

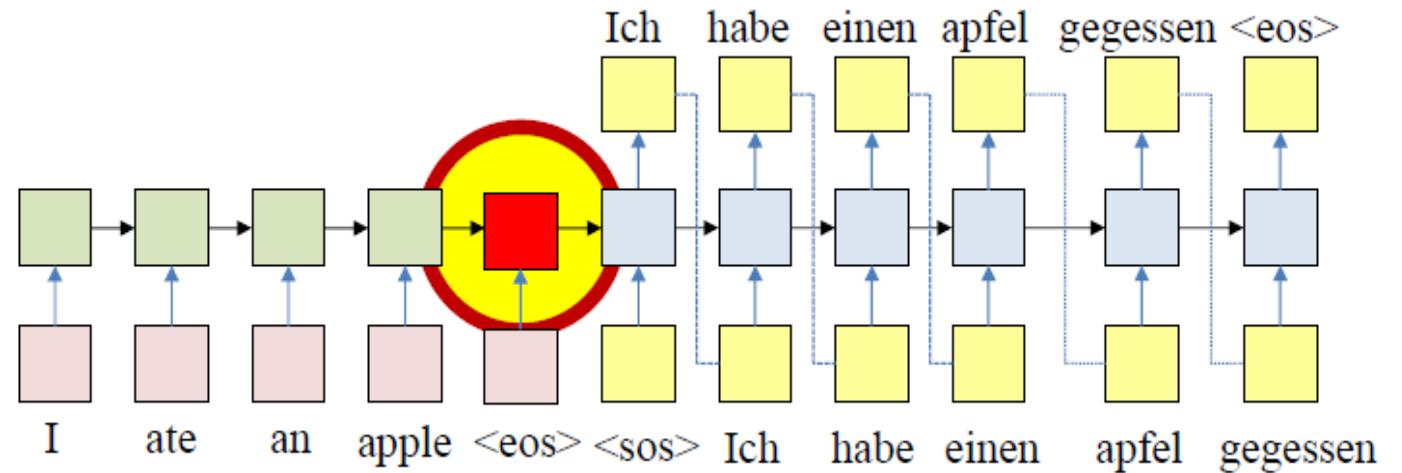- Over-predict Rainfall Event occurrences
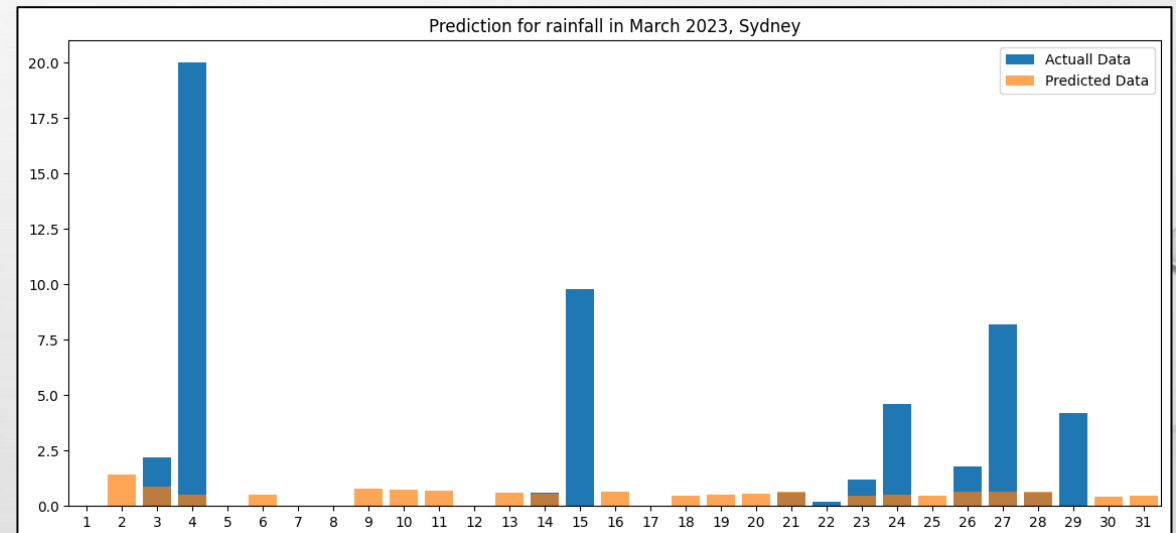
LIMITATION

Model prediction only relies on last hidden stage

Information in all previous hidden stages not utilized

Information might not be retained through sequence length

FUTURE WORK

Model prediction only relies on last hidden stage

Information in all previous hidden stages not utilized

Information might not be retained through sequence length

Modified LSTM
- Incorporate attention
- Multiply zero-inflated input with input gate sigmoid unit

Attention Mechanism

Transformer

THANK YOU
Q&A